THREE-DIMENSIONAL-ENERGY MINIMIZED MODELS FOR CALF SKIN TYPE I COLLAGEN TRIPLE HELIX AND MICROFIBRIL: I. THE TRIPLE HELICAL MODELS

Abstract

Molecular modeling methods were applied to the development of three-dimensional structures of Type I collagen triple helices. The initial model was based on a triple helix of $(Gly\text{-}Pro\text{-}Hyp)_{12}$. The amino acid sequences of the alpha 1 and 2 chains of calf skin Type I collagen were substituted into the $(Gly\text{-}Pro\text{-}Hyp)_{12}$ model. Each Type I structure was energy minimized to form all possible stabilizing interactions within the protein. Specific interaction sites which may contribute to the stability of the collagen molecule can be identified in the triple helical models. The models may be used to target potential domains in collagen that contain sites for crosslinking reactions with chromium and other agents; determine what constraints on size, shape and molecular characteristics are likely to describe the optimum crosslinking agents; and suggest the application of which analytical and spectroscopic techniques will give the most rigorous test of the computer models.

Introduction

Effective tanning of limed hide is essential for the production of quality leather. Compared to the native state of calf skin, leather has the properties of strength and resistance to biological organisims which attack and destroy hide. At present, Cr(III) is highly effective as a tanning agent. Although other tanning agents do exist (i.e., vegetable tannins, synthetic organic crosslinking agents and other mineral agents), trivalent chromium is the most effective in producing quality leather from calf skin[1]. There is concern that nonhazardous Cr(III) in solid leather waste may be oxidized to Cr(VI)[2,3]. Cr(VI) oxide (CrO_3) is soluble in water and is poisonous[4]. Based on these properties, Cr(VI) is hazardous to humans and to the environment. An alternative tanning agent for the production of leather may be desirable. This new agent must have equal or superior tanning properties compared to that of Cr(III), without undesirable attributes.

Experimental procedures for identifying alternative agents are based on trial and error testing of numerous compounds or mixtures of known tanning agents[5-8]. If a set of possible compounds were identified to have tanning properties, the next step would be to synthesize the derivatives of these compounds for obtaining more reactive species. Metals such as aluminum and zirconium have been tested as ligands, but are less effective than Cr(III)[1].

Overall, the identification of new tanning agents is based on experimental trials of compounds which are similar in property to known ligands. In order to design a new and specific ligand, the inter- and intra-molecular interactions of collagen must be understood. Moreover, if the structure of Type I collagen were known, possible reaction sites of tanning ligands could be identified. By carrying out structure-function studies based on protein-ligand interactions it may be possible to determine the chemical and geometric properties of known tanning ligands. The characterization of these relationships would lead to a more efficient and effective procedure for designing new or improved tanning agents. The key barrier to this approach is that the three-dimensional structure of Type I collagen is not known. This study describes the application of molecular modeling to the study of the three-dimensional structure of collagen.

Molecular modeling techniques are used in order to better understand the structure-function relationships of collagen and collagen interactions. The goals of this study are to construct a three-dimensional model for the Type I collagen triple helix and to identify the effects of tanning agents (i.e., Cr(III)) in modifying the structure and interactions of native collagen. Inter- and intra-polypeptide interactions between sidechains and/or backbone groups will be described. Identification of important sites for ligand interactions will also be discussed. In addition, procedures for further study and refinement of the initial models of the Type I collagen triple helix will be presented.

Background Review on the Structure of Collagen

The Collagen Polypeptide Sequence

The amino acid sequences of the alpha 1[9] and 2[10,11] chains of Type I collagen are known. Excluding the extra-helical telopeptide chains, each chain consists of 1014 amino acid residues[9-11]. Each polypeptide sequence contains the consensus tripetide sequence $(Gly\text{-}X\text{-}Y)_n$ (n is equal to 338 per chain), where glycine is always in the first position and "X" and "Y" are the corresponding amino acids which vary within each chain[12]. The triple helical region of each polypeptide chain contains 33% glycine and about 25% imino acids, proline and hydroxyproline[12]. The significance of these amino acid residues (Gly, Pro and Hyp) in the sequence of collagen has resulted in the use of synthetic polypeptide models composed of the consensus sequence Gly-Pro-Pro or Gly-Pro-Hyp for studying the structure-function relationships of collagen[13-17]

The Triple Helix

Based on earlier model building[18,19] and X-ray diffraction studies[20], the structure of collagen is known to be a triple helix of three polypeptide chains. Each chain has a left handed helical twist, taking approximately 3.3 amino acid residues to make one complete (360 degrees) rotation about the polypeptide helical long axis[20]. The interaction of three collagen polypeptide chains forms a right-handed triple helix. It takes 27-29 amino acid residues, depending on the models used, for a single polypeptide chain to make one complete rotation about the triple helical long axis. Within the collagen molecule, each polypeptide chain has two hydrogen bonds per Gly-X-Y consensus tripeptide sequence. These hydrogen bonds are formed between the backbone carbonyl oxygen (X-position) of a single tripeptide sequence in one polypeptide chain and the backbone amide hydrogen

(Gly) of another tripeptide sequence in an adjacent polypeptide chain. Thus, each tripeptide of Gly-X-Y participates in two hydrogen bond interactions.

Methods

Computer Hardware and Software

Molecular modeling studies were conducted on a computer system consisting of two Silicon Graphics 4D/25 processors, integrated to both a Silicon Graphics and an Evans & Sutherland PS390 high resolution graphics workstation. This color graphics system allows for detailed visualization and real time manipulation of chemical structures. SYBYL (v.5.32, from Tripos associates, Inc., 1990[21]), was the primary software used for molecular modeling. SYBYL contains computational tools for protein construction, manipulation and energy minimization.

Energy Minimization

Energy minimization of protein structures was accomplished by the molecular mechanics method (see Ref. 22 and 23 for review). This method contains a set of force-fields which have been parameterized to reproduce experimentally determined properties (i.e., bond length and bond angles) of known molecules. Thus, the purpose of the molecular mechanics methods is to predict the structural and energetic properties of molecules which can not be easily measured experimentally. The success of this method is based on evidence that the force-fields are transferrable between different molecules or sets of molecules. Specific forces and interactions which stabilize the three-dimensional structure of a given protein are represented as separate potential energy values. The total potential energy value, E_{tot}, represents the sum of these separate energy terms and defines a protein's three-dimensional structure. For a specific protein, it is assumed that the native structure is the structure of lowest total potential energy. Hence, the energy of a protein model is minimized computationally in order to lower its E_{tot} value. The SYBYL OPTIMIZATION MAXIMIN2 option includes several energy minimization algorithms using either the Tripos or AMBER force-fields[21]. For the collagen models, the conjugate gradient minimization method with the AMBER force-field was used[21-23].

Construction of the Type I Triple Helical Models
The (GLY-PRO-HYP)$_{12}$ Triple Helix

Synthetic polypeptide chains of (Gly-Pro-Hyp)$_n$, where n = 10, have been shown to form collagen-like triple helical structures[13-17]. Therefore, construction of the three-dimensional structures for the Type I triple helix was based on the triple helical model for (Gly-Pro-Hyp)$_{12}$[24]. This model was constructed from three separate polypeptide chains of (Gly-Pro-Hyp)$_{12}$ where each chain was assigned the helical conformation known for collagen. The chains were aligned on the graphic display to simulate the collagen triple helix. This collagen model was then energy minimized to derive all possible stabilizing interactions[24]. When energy minimization of the (Gly-Pro-Hyp)$_{12}$ triple helix was complete, additional copies of this structure were made[21] and used as a basis for amino acid substitutions. Pro and Hyp residues were replaced by the appropriate amino acid residues from the X and Y positions of the alpha 1 and 2 chains of Type I collagen.

The Type I Triple Helical Models

Five Type I triple helical models representing different regions of the collagen molecule were constructed as shown in the alignment scheme of Figure 1 for the packing of adjacent triple helical collagen molecules. The alignment scheme in Fig. 1 is based on the 'Smith' microfibril model, containing five laterally packed collagen triple helices[25]. Thus, the Type I triple helical models were also used to study inter-helical interactions involved in stabilizing collagen packing[24]. Each collagen molecule contains the amino acid sequences for two alpha 1 and one alpha 2 chains. Figure 1 shows the alignment pattern of a complete collagen molecule (labeled 1) and the corresponding regions of four adjacent molecules (labeled 2-5) as proposed by Smith[25]. In this model, each collagen molecule with respect to the adjacent molecule is staggered laterally, along their helical axis, by 234 residues or 1 D spacing (labeled 'D-spacing' in Fig. 1). In the 'Smith' microfibril (Fig. 1), label 'a' refers to the 'overlap' region containing five triple-helical structures and label 'b' refers to the 'gap' region containing four helical structures. Electron microscopy studies[12,24] have shown that there are no direct end-to-end interactions between adjacent collagen molecules along the same vertical axis; therefore, 'gap' regions ('b' in Fig. 1) exist where there are no helical collagen molecules. Figure 2 contains the specific sequences for the construction of the five Type I models based on the 'Smith' model[25]. The first triple helical model was substituted with the Type I polypeptide sequences of two alpha 1 and one alpha 2 chains containing tripeptides 17-28 (Fig. 2). The other four models were constructed similarly using tripeptides 95-106, 173-184, 251-262 and 329-338. The amino terminal residue of each model (i.e., labeled COLLAGEN 1-5, Fig. 2) is 234 residues or 1D spacing (78 tripeptides) further along the sequence than the previous model. Tripeptide 338 represents the actual carboxyl terminus of the triple helical region of Type I collagen, thus the model (labeled COLLAGEN 2, Fig. 2) beginning with tripeptide 329 contains only 10 tripeptides and terminates in the 'gap' region[25] (gap regions are labeled as 0 in Fig. 2).

Results and Discussion

Energetic Evaluation of the Type I Collagen Triple Helices

Table I gives the computed total energy (E_{tot}) for each of the five Type I triple helical models before minimization (E_{toti}), after energy minimization (E_{totf}) and the energy difference (E_{totd}) between the two previous values which is taken as: $(E_{totd}) = (E_{totf}) - (E_{toti})$. The energies of the structures after minimization are lower (more favorable) than before minimization as indicated in the E_{totd} values for collagen 1-5 (Table I). This indicates that stabilizing interactions were formed.

The computed total potential energy after minimization, E_{totf}, for the $(Gly\text{-}Pro\text{-}Hyp)_{12}$ triple helix was -94.0 kcal/mole. Compared to the E_{totf} values for the Type I models (collagen 1-5 in Table I), the $(Gly\text{-}Pro\text{-}Hyp)_{12}$ model has a higher total potential energy. Although a direct comparison of the E_{totf} values can not be made for structures containing different amino acid sequences, the large difference in E_{totf} values when comparing each Type I model to the $(Gly\text{-}Pro\text{-}Hyp)_{f12}$ model indicates that sidechain interactions due to the native collagen sequences are very important for stabilizing the triple helix. Although models based on a repeat of the tripeptide sequence (Gly-Pro-Hyp) are useful for studying certain structure-function relationships of collagen, models containing the actual Type I sequences represent a realistic and more accurate system.

FIG. 1. — Sequence alignment of calf skin Type I collagen into the packing arrangement proposed by Smith[25] containing five collagen triple helices polarized in the same direction from amino (top) to carboxyl (bottom) end within a subfibrillar unit. The complete triple helical region is seen for the first collagen molecule (labled 1). Collagen 2-5 (labeled 2-5) consist of the corresponding triple helical regions which are packed parallel and adjacent to Collagen 1 in the Smith microfibril model[25]. Each respective collagen is aligned such that there is a stagger, laterally, between each triple helix referred to as 1D spacing (labeled 'D-spacing', D=234 amino acid residues per chain). The regions seen as 'a' and 'b' represent the 'overlap' and 'gap' regions, respectively.

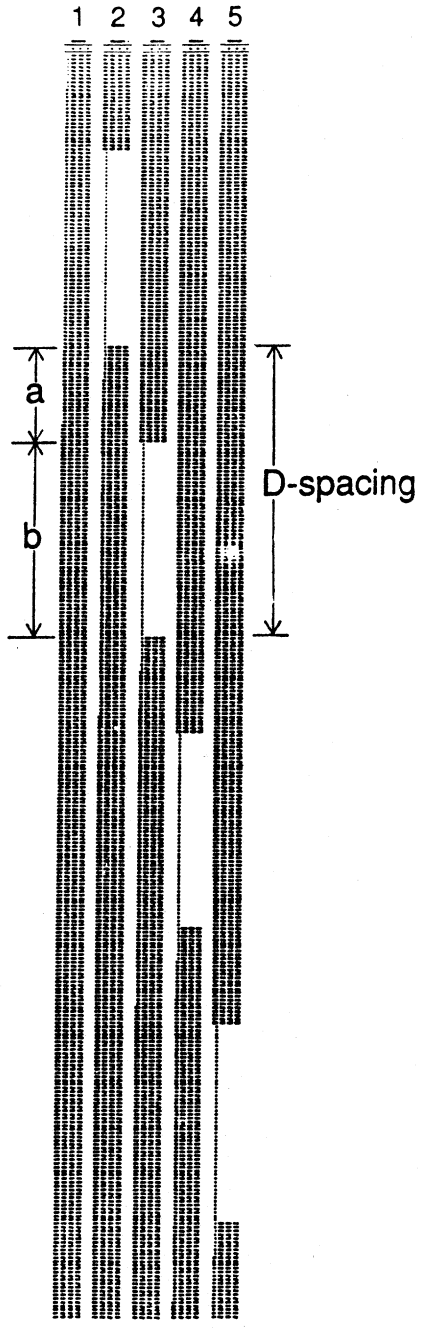


FIG. 2. — Seen here are the specific triple helical regions based on the alignment scheme in Fig. 1 which will be used to construct five different energy minimized Type I models. A1 and A2 refer to the alpha 1 and alpha 2 polypeptide chains of collagen, respectively. In each collagen triple helix, there are two alpha 1 and one alpha 2 chains, respectively. Tripeptide 338 in the COLLAGEN 2 sequences represents the actual carboxyl terminus of a single Type I molecule and this region contains part of the 'gap' region (labeled 0). The alignments shown here have a length of 36 amino acid residues or 12 tripeptides per collagen molecule all polarized in the same direction with respect to the COLLAGEN 2 triple helix. The amino terminal portions of the five helical regions have a high density of charged sidechains (17-21, 329-333, 251-255, 173-177 and 95-99 for COLLAGEN 1-5, respectively) and the carboxyl terminal portions have a high density of non-polar sidechains (25-28, 334-338, 259-262, 180-184 and 106 for COLLAGEN 1-5, respectively). Single letter codes where the 'upper case' letters represent naturally occuring amino acids and the 'lower case' letters represent the corresponding modified 'hydroxylated' amino acids are used. Naturally occuring amino acids are: A = alanine; D = aspartate; C = cysteine; E = glutamate; F = phenylalanine; G = glycine; H = histidine; I = isoleucine; K = lysine; L = leucine; M = methionine; N = asparagine; P = proline; Q = glutamine; R = arginine; S = serine; T = threonine; V = valine; Y = tyrosine.

SEQUENCES USED IN THE TYPE I TRIPLE HELICAL MODELS

COLLAGEN 1			COLLAGEN 2			COLLAGEN 3			COLLAGEN 4			COLLAGEN 5		
A1	A2	A1	A1	A2	A1	A1	A2	A1	A1	A2	A1	A1	A2	A1
17 GKN	GKA	GKN	329 GPp	GPA	GPp	251 GPA	GFV	GPA	173 GPR	GSR	GPR	95 GAA	GPS	GAA
18 GDD	GED	GDD	330 GPR	GIR	GPR	252 GEK	GEp	GEK	174 GAN	GPS	GAN	96 GEE	GEE	GEE
19 GEA	GHp	GEA	331 GRT	GSQ	GRT	253 GAp	GPS	GAp	175 GAp	GPp	GAp	97 GKR	GKR	GKR
20 GKP	GKP	GKP	332 GDA	GSQ	GDA	254 GAD	GEp	GAD	176 GND	GPD	GND	98 GAR	GST	GAR
21 GRp	GRp	GRp	333 GPA	GPA	GPA	255 GPA	GTA	GPA	177 GAK	GNK	GAK	99 GEp	GEI	GEp
22 GER	GER	GER	334 GPp	GPp	GPp	256 GAp	GPp	GAp	178 GDA	GEp	GDA	100 GPS	GPA	GPS
23 GPp	GVP	GPp	335 GPp	GPp	GPp	257 GTP	GTT	GTP	179 GAp	GVV	GAp	101 GLp	GPp	GLp
24 GPQ	GPQ	GPQ	336 GPp	GPp	GPp	258 GPQ	GPQ	GPQ	180 GAp	GAp	GAp	102 GPp	GPp	GPp
25 GAR	GAR	GAR	337 GPp	GPp	GPp	259 GIA	GLL	GIA	181 GSQ	GTA	GSQ	103 GER	GLR	GER
26 GLp	GFP	GLp	338 GPP	GPP	GPP	260 GQR	GAp	GQR	182 GAp	GPA	GAp	104 GGp	GNp	GGp
27 GTA	GTp	GTA	0			261 GVV	GFL	GVV	183 GLQ	GPS	GLQ	105 GSR	GSR	GSR
28 GLp	GLp	GLp	0			262 GLp	GLp	GLp	184 GMp	GIp	GMp	106 GFp	GLp	GFp

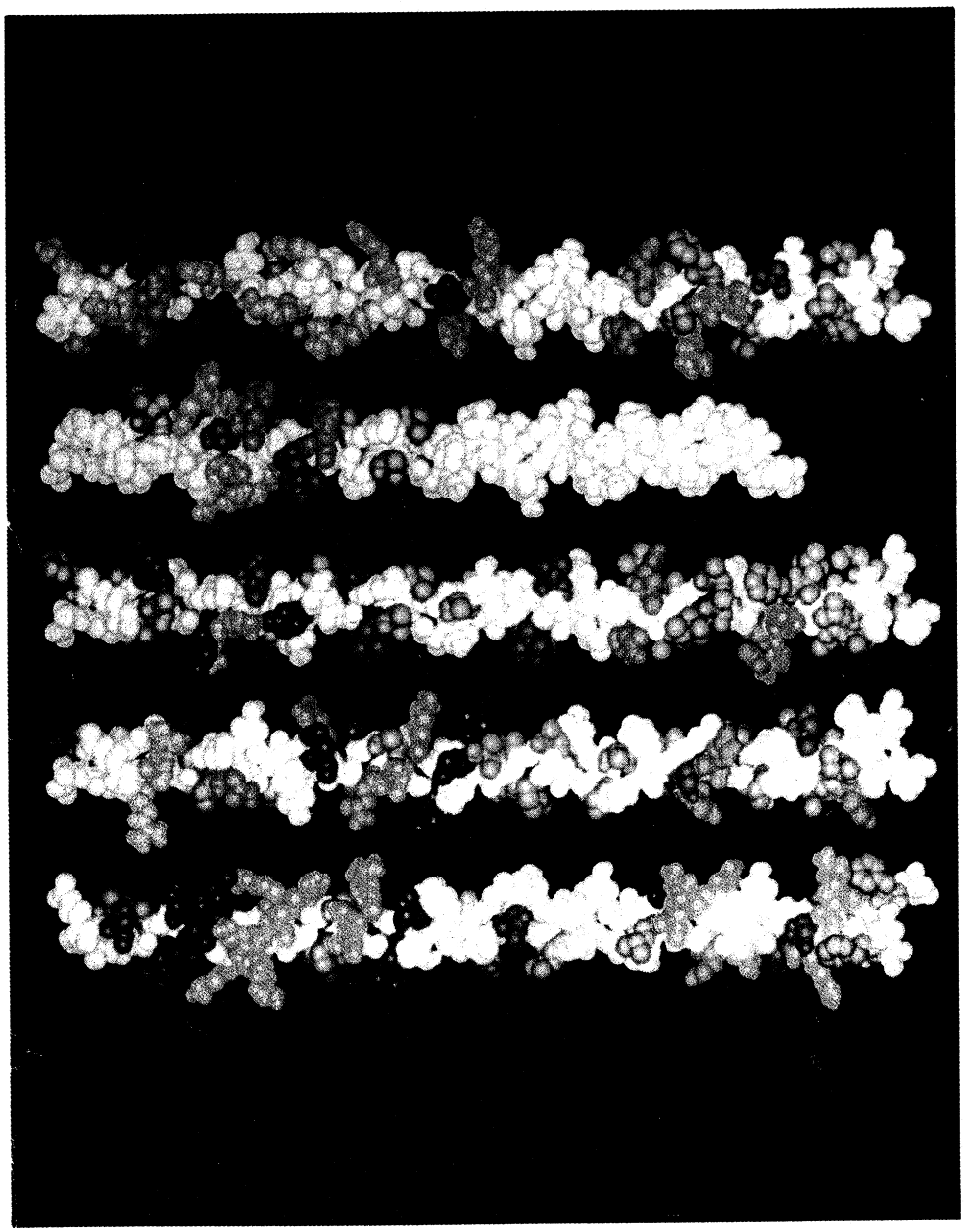
TABLE I

Computer Total Potential Energies for the Triple Helical Models of Type I Collagen

Energy Term[a]	Collagen[b] 1	Collagen[b] 2	Collagen[b] 3	Collagen[b] 4	Collagen[b] 5
E_{toti}	- 682.954	-354.350	-873.935	537.292	- 697.925
E_{totf}	-1114.215	-404.215	-969.774	-1016.125	-1235.240
E_{totd}	- 431.261	- 50.482	- 95.839	-1553.417	- 537.315

[a]Energies were computed in kcal/mole using the AMBER force fields with united-atoms[21,24,26] according to the potential energy function in equation 1 as described in Methods. The potential total energies for: before minimization (E_{toti}), after minimization (E_{totd}) are computed by taking the difference: (E_{totf})-(E_{toti}). Although the following energy values are not shown in this table, the individual energy terms are defined as follows: E_{bs} is the sum of energies arising from bond stretching or compression beyond the optimum bond length; E_{ab} is the sum of energies for angles which are distorted from their optimum values; E_{tor} is the sum of the torisonal energies which arise from rotations about each respective dihedral angle; E_{op} is the sum of energies for the bending of planar atoms out of the plane; E_{vdw} and E_e are the sum of energies due to non-bonded van der Waals and electrostatic interactions, respectively; E_{14vdw} and E_{14e} are the sum of the van der Waals and electrostatic interaction energies, respectively for atoms connected by three bonds and E_{hb} is the sum of energies due to hydrogen bond interactions[21,24]. The non-bonded E_{vdw} and E_e interaction terms were given a cutoff distance of 8 angstroms in which all the interacting atom pairs separated by a distance greater than 8 angstroms are not accounted for in the energy calculations. Solvent effects of water were not included explicitly, but were accounted for implicitly in a dielectric function (R_{ij} + 1) where R_{ij} is the distance between two separated atoms[21,24]

FIG. 3. — Space-filling models (Fig. 3) of the energy minimized triple helical structures for Type I collagen where tripeptide 338 (i.e., the carboxyl terminus) is the top-end of the second structure (counting from the left). All the three-dimensional triple helical structures are polarized in the same direction and show a variety of intra- and inter-polypeptide sidechain-sidechain and sidechain-backbone interactions. In this figure, colors are used to distinguish different chemical properties of the amino acid residues. In the three-dimensional models, it is clear that the amino acid sequences of Type I collagen define several interesting functional domains. For example, the top regions of all the triple helical models contain many non-polar sidechains (colored orange) and the bottom regions contain mainly charged acidic and basic sidechains (colored red and blue, respecctively). Asparagine and glutamine residues are green; threonine and serine residues are purple.



Type I Collagen Triple Helices

Figure 3 shows the space-filling models of the energy minimized triple helical structures for Type I collagen where tripeptide 338 (i.e., the carboxyl terminus) is located at the top-end of the second structure (counting from the left of Fig. 3). The five triple helical structures are polarized in the same direction and show a variety of intra- and inter-polypeptide sidechain-sidechain and sidechain-backbone interactions. Colors are used to distinguish different chemical properties of the amino acid residues. In the three-dimensional models, it is clear that the amino acid sequences of Type I collagen define several interesting functional domains. For example, the top regions of each of the triple helical models contain many non-polar sidechains (colored orange) and the bottom regions contain mainly charged acidic and basic sidechains (colored red and blue, respectively). The clustering of amino acid residues with similar sidechain properties, based on the 'Smith' model[25], is very important for the specific interactions between adjacent collagen molecules during fibril formation[24,26]. Each structure was examined graphically in order to describe possible stabilizing intra- and inter-polypeptide interactions. Possible structure-function relationships established from detailed interactive studies are described below.

Structural Evaluation of Stabilizing Sidechain Interactions Found for the Triple Helical Models of Type I Collagen

Structure-Function Relationships Found for Arginine

The positively charged sidechain of arginine is found in both the 'X' and 'Y' positions of the consensus sequence Gly-X-Y in Type I collagen[12]. In addition to its ability to form charged interactions, the sidechain guanidinium group contains five amino protons which are potential hydrogen bond donors. These properties of arginine implicates that it may be important in stabilizing collagen structures.

The triple helical regions of the alpha 1 and 2 polypeptide chains contain 50 and 55 arginines respectively out of 1014 amino acid residues per chain[9-11]. In the energy minimized triple helical models, Arg at the 'Y' positions of Gly-X-Y have their sidechains in an extended conformation, facing outwards from the helical long axis of the collagen model while arginines in the 'X' positions are folded against the surface of the collagen chains. The difference in sidechain conformations for arginines in the 'X' or 'Y' positions are due to the different conformational constraints imposed for each position in the three-dimensional structure of the collagen polypeptide chain. It can be postulated that Arg in the 'X' position functions to stabilize the triple helical structure by reducing the conformational flexibility of the polypeptide chain. Arginine at the 'Y' positions may function in inter-helical packing interactions where the sidechains crosslink to functional groups (i.e., acidic and polar) of adjacent helices. In calf skin Type I collagen, there are 42/50 and 43/55 arginines (per polypeptide chain) found in the 'Y' positions for the alpha 1 and alpha 2 chains, respectively[9-11]. The high ratio of arginines found as Gly-X-Arg indicate the importance of arginine in collagen packing.

Non-Polar Interactions in the Triple Helical Models

Examination of the complete amino acid sequence for calf skin Type I collagen shows that the non-polar sidechains are localized into clusters along the helical long axis of the collagen polypeptide chains (see Fig. 2 in the region for COLLAGEN 1-5 corresponding to the tri-peptides 25-28 of COLLAGEN 1; see Fig. 3 for regions colored orange). Hydro-

phobic forces which contribute to the packing of collagen molecules result from the properties of these non-polar sidechains (leucine, isoleucine, valine, phenylalanine, methionine, proline and alanine). Figure 3 shows that the non-polar sidechains are directed outwards, away from the center of the triple helical models. If solvent (water) were explicitly included in the models, the positions of these non-polar sidechains would result in a very unfavorable free-energy of hydration. Similar to globular proteins where non-polar sidechains are found within the interiors, away from water, the non-polar surfaces of collagen triple helices would tend to be packed together (due to the hydrophobic effect[27]), shielding themselves from the hydrophilic environment. Hence, the non-polar sidechains which are found in clusters along the Type I sequences contribute to the inter-helical packing forces between different collagen molecules.

Asparagine-Aspartic Acid Interactions in Stabilizing the Triple Helical Models

A set of adjacently positioned aparagine-aspartic acid residues form inter-polypeptide stabilizing interactions in the three-dimensional models containing the tripeptide regions 173-184. This localized region contains three pairs of aparagine-aspartic acid interactions which may be important in the stabilization of the triple helix of collagen. Furthermore, if the aspargine sidechains were to be hydrolyzed into aspartic acids, there would be an increased number of potential Cr(III) binding sites[26].

Interactions Between Lysine and Glutamate Observed in the Three-Dimensional Collagen Models

Several stabilizing interactions due to the charged sidechains of Lys and Glu are seen in the tripeptide region 95-106. Both intra- and inter-polypeptide interactions of Lys and Glu are observed. The intra-polypeptide Lys/Glu interactions have also been accounted for in a recent computational study for collagen models[28], but this study did not include inter-polypeptide interactions. Although specific Lys/Glu interactions have been observed, there are probably other stabilizing configurations for the above pairing of basic and acidic sidechains. The sidechain of lysine is highly flexible within proteins[29] which would likely reduce the specificity or strength of the above Lys/Glu interactions. Furthermore, when a collagen triple helix interacts with adjacent collagen molecules, these interactions would be modified in accordance with the new protein environment. It is this reduced-specificity and potency of the acid/base sidechain interactions in native collagen which allows for the proposal of an initial hypothesis explaning why Cr(III) is an effective tanning agent for leather processing.

A Hypothetical Model for the Interaction of Tanning Agents Pertaining to the Lysine/Glutamate Active Sites in the Collagen Triple Helix

In the above description of Lys/Glu interactions within the collagen triple helical model, some implications can be proposed for the activity of tanning agents. The triple helical region of Type I collagen contains approximately 30 tripeptide sequences of the type Gly-X-Lys and 12 sequences of the type Gly-Lys-Y[9-11]. Each lysine rich region is in close proximity to acidic groups such as aspartate and glutamate. This type of configuration for the acidic and basic sidechains optimizes salt-link formations which would stabilize both the triple helix and triple helical interactions. In order for tanning agents which interact at the acidic sites (i.e., Cr(III)) to compete in stabilizing the collagen matrix, these

modifying agents must have a higher affinity/activity for the acidic groups than that of the native collagen groups which are originally complexed with glutamate or aspartate. In the case of Lys interactions with Glu or Asp, it can be hypothesized that Cr(III) competes better than other known tanning agents for the acidic sites. Furthermore, the lysine sidechains which are no longer bound to their acidic counterparts are available for cross-linking by other modifying agents which interact at the basic sites. Hence, according to the above hypothesis, Cr(III) is a special tanning agent which stabilizes collagen structures and also creates possible new sites for further stabilization by other tanning agents. In order to design better tanning agents than Cr(III), chemical reagents which can better compete for the acidic sites must be identified. Under the conditions that Cr(III) is an effective tanning agent (such as the pH and molar conc.[1]), its geometrical (i.e., size, shape) and reactive (i.e., binding affinity[1]) properties must be determined. Computationally, the 'Thermodynamic Cycle-Peturbation Methods[30-32] could be applied in order to calculate and compare the relative free energies of binding for different tanning ligands. Once the basis for the high reactivity of chromium is understood, modified or new chemical ligands could be designed. To determine which geometric parameters result in the most effective distribution of the tanning agent throughout the collagen matrix, the three dimensional structure of collagen packing and dynamics should be studied. Both the geometric and chemical properties which define a specific and ideal tanning ligand must be incorporated into the design of the new ligand. The structural refinement studies of collagen interaction are in progress.

Conclusion

Type I collagen triple helical models have been constructed based on the energy minimized model of $(Gly\text{-}Pro\text{-}Hyp)_{12}$[24]. Energy minimization of these three-dimensional models allows for the sidechains to form stabilizing interactions. Using computer graphics display, the Type I triple helices can be observed three-dimensionally. In addition, color enhancement of specific residues in the models allows for quick and effective examination of detailed structure-function relationships. In the three-dimensional models, possible sidechain conformations can be studied directly. Important sidechain interactions can be studied and related to interactions with other proteins or tanning ligands. Any triple helical region in the linear amino acid sequence of Type I collagen can be incorporated into the models and studied[24]. Using data from published experimental results such as the specific formation of sidechain crosslinks or results pertaining to ligands which bind to specific regions of collagen, molecular models representing different segments of Type I collagen can be studied and compared in order to establish subfibrillar or fibrillar structures which correlate with the above experimental data. This procedure allows for a more precise interpretation of experimental data. In addition, molecular modeling combined with experiments allow for more effective and efficient designing of future experiments. Overall, molecular modeling of collagen increases the precision of studying protein-ligand interactions and increases the efficiency of obtaining or deriving specific geometric information on which new or improved tanning agents can be based.

Further study of the collagen models will involve the identification of all the possible stabilizing sidechain-sidechain and sidechain-backbone interactions (i.e., such as specificially formed hydrogen bonds, charge-charge and nonpolar interactions). The Type I models will then undergo structural modification in order to maximize all possible interactions. However, it is clear that the Type I triple helical models described here provide a reasonable three-dimensional view of possible collagen interactions.

Acknowledgement

The authors would like to thank Drs. C.E. Kung and A.J. Miller for helpful suggestions and review of this manuscript.

References

1. Bienkiewicz, K.J., In "Physical Chemistry of Leather Making", R.E. Krieger Pub. Co., Florida, pp. 324-352, (1983).
2. Silva, S., Agricultural use of leather working residues, UNIC, Milano, Italy, (1989).
3. Rutland, F.H., Kallenberger, W.E., Menden, E.E., and Nazariv, C.L., Leather Industries of American Bulletin, #89-5, Leather Industries of America, Washington, D.C. (1989).
4. Cotton, F.A. and Wilkinson, G., In "Advanced Inorganic Chemistry", F.A. Cotton, G. Wilkinson, eds, John Wiley & Sons, New York, pp. 719-736, (1980).
5. Deme, I., Proceedings of the XXth Congress of the IULTCS, #S-4, (1989).
6. He, X., and Guo, Z., Proceedings of the XXth Congress of the IULTCS, #S-11, (1989).
7. Heidemann, E., Tonigold, L., and Hein, A., Proceedings of the XXth Congress of the IULTCS, #S-12, (1989).
8. Pauckner, W., Proceedings of the XXth Congress of the IULTCS, #S-14, (1989).
9. Dolz, R., and E., and Heidemann, E., *Biopolymers,* **25,** 1069-1080, (1986).
10. Boedtker, H., Finer, M., and AHO, S., *Ann. N.Y. Acad. Sci.,* **460,** p. 85, (1985).
11. Heidemann, E., pesonal communications.
12. Ramachandran, G.N., and Ramakrishnan, C., In "Biochemistry of Collagen", G.N. Ramachandran, ed., Plenum Press, New York, pp. 45-84, (1976).
13. Berg, R.A., and Prockop, D.J., *Biochem. Biophy. Res. Comm.,* **52,** 115-120, (1973).
14. Sakakibara, S., Kishida, Y., Okuyama, K., Tanaka, N., Ashida, T., and Kakudo, M., *J. Mol. Biol.,* **65,** 371-373, (1972).
15. Olsen, B.R., Berg., R.A., Sakakibara, S., Kishida, Y., and Prockop, D.J., *J. Mol. Biol.,* **57,** 589-595, (1971).
16. Inouye, K., Hayakawa, T., Kobayashi, Y., Sugeta, H., and Kyogoku, Y., In *"Peptide Chemistry"*, N. Izumiya, ed., Protein Research Foundation, Osaka, pp. 121-126, (1985).
17. Bhatnager, R.S., Pattabiraman, N., Sorensen, K.R., Langridge, R., MacElroy, R.D., and Renugopalakrishnan, V., *J. Biomolecular Structure & Dynamics,* **6,** 223-233, (1988).
18. Rich, A., and Crick, F.H.C., *J. Mol. Biol.,* **3,** 483-506, (1961).
19. Ramachandran, G.N., Sasisekharan, V., and Thathachari, Y.T., In *"Collagen"*, N. Ramanathan, ed., Interscience Publishers, New York, pp. 81-116, (1962).
20. Ramachandran, G.N. *Am. Leather Chemists' Assoc.,* **63,** 160-191, (1968).
21. SYBYL, Macromolecular modeling software, version 5.32, TRIPOS Associates, Inc., (1990).
22. Allinger, N.L., *Adv. Phys. Org. Chem.,* **13,** 1-82, (1976).
23. Boyd, D.B., and Lipkowitz, K.B., *J. Chem. Educ.,* **59,** 269-274, (1982).
24. Chen, J.M., Kung, C.E., Feairheller, S.H., and Brown, E.M., *J. Protein Chem.,* **10**(5), 535-553 (1991).
25. Smith, J.W., *Nature,* **219,** 157-158, (1968).
26. Chen, J.M., Feairheller, S.H., and Brown, E.M., *J. Am. Leather Chemists' Assoc.,* (1991).
27. Kauzmann, W., *Adv. Protein Chem.,* **14,** 1-63, (1959).
28. Katz, E.P., and David, C.W., *Biopolymers,* **29,** 791-798, (1990).